DOCUMENT RESUME

ED 219 427 TM 820 442

AUTHOR Lohmoeller, Jan-Bernd
TITLE An Overview of Latent Variables Path Analysis.
PUB DATE Mar 82
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Covariance; Comparative Analysis; Estimation (Mathematics); Foreign Countries; *Least Squares Statistics; *Mathematical Models; *Maximum Likelihood Statistics; *Path Analysis
IDENTIFIERS *Latent Variables; *LISREL Computer Program; West Germany

ABSTRACT

Latent Variables Path (LVP) models have been found to be a helpful device for the formulation and validation of complex theories in educational sciences. The estimation of parameters for LVP models can use the Maximum Likelihood (ML) or the Partial Least Squares (PLS) method. This paper focusses on the PLS method introduced by H. Wold, and compares it to the LISREL method introduced by K.G. Joereskog. The differences in the results are explained by the differences in the statistical assumptions, and some clues to the choice of ML or PLS are given. Several select applications show the range where only PLS estimation is applicable. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED219427

An overview of Latent Variables Path Analysis

Jan-Bernd Lohmoeller

University of the Federal Armed Forces Munich

Federal Republic of Germany

Paper presented at the Annual Meeting 1982

American Educational Research Association

New York, March 19 to 23, 1982

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✗ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J.B. Lohmoeller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Summary: Latent Variables Path (LVP) models have been found to be a helpful device for the formulation and validation of complex theories in educational sciences. The estimation of parameters for LVP models can use the Maximum Likelihood (ML) or the Partial Least Squares (PLS) method. This paper focusses on the PLS method introduced by H.Wold, and compares it to the LISREL method introduced by K.G.Joereskog. The differences in the results are explained by the differences in the statistical assumptions, and some clues to the choice of ML or PLS are given. Several select applications show the range where only PLS estimation is applicable.

Section	Page	Content:
Sc.1	2	Introduction
Sc.2	4	A two-block model, estimated by PLS and ML
Sc.3	8	The statistical properties: consistency
Sc.4	10	Covariance structure, data structure models
Sc.5	15	Some selected applications of PLS
Sc.6	16	Conclusions

TM 820 442

An overview of Latent Variables Path Analysis.

1. Introduction.

Latent Variables Path Analysis (LVPA) considers a set of latent variables (LVs), measured by a set of manifest variables (MVs) or other latent variables, and related to each other by linear equations. The LVP model as well as the first viable estimation technique was developed by Karl Joereskog (1973, 1978). His LISREL method (Joereskog & Soerbom 1981) has found wide respect for its mathematical, statistical, and substantive clarity, but less widespread enthusiasm for the computing time his LISREL program needs. Bentler and Weeks (1980) apply GLS (generalized least squares) to LVP models. An alternative approach to LVPA will be presented in this session, called the Partial Least Squares (PLS) technique of LVPA, or the of Soft Modeling, as its author Herman Wold (1975 onwards) prefers to call it.

A systematic treatment of LVP analysis can be found in the excellent review by Bentler (1980 a,b), as well in the paper by Jack McArdle and McDonald (1981). The reader who is interested in the summary of the ML methods is referred to these papers. The PLS approachs is only mentioned, not treated in these papers, so this will be done here.

The PLS approach is implemented in a computer program called LVPLS (Lohmoeller 1981b), where the first part of the name is related to the type of the models to be estimated by this program, namely LVP models, and the second part of the name is related to the estimation technique, the PLS estimation method. The program is written in standard Fortran, and is tested on eight different brands of computers. It is available at nominal costs.

The presentation of this paper will focus on the essential differences between the PLS and ML estimation technique, not so much on the technical differences between the programs available at time.

As a first example of LVP modeling, two graphs from the Review of Educational Research will be used. The first example, taken from Shavelson and Stern (1981) shows the factors contributing to teachers' pedagogical judgements and decisions, where some of the factors themselves are interrelated by arrows. We may assume that the arrows stand for causal or at least predictive relationships, and the factors can be called causes or at least predictors. This model is not specified in such details that it can be used immediately to design a questionaire or research project. It is merely a very rough blueprint, serving to relate all hypotheses under question, and to check the consistency and completeness of the set of hypotheses.

The second example, taken from the same issue of Review of Education, is explicitely related to the statistical method, here cited as structural modelling. Each one of the six blocks contains the name of a (uni- or multidimensional) latent variable, and several manifest variables (indicators) are listed within parentheses. Tough no confrontation of this model as a whole with empirical data is reported, some of the studies Glasman and Biniaminov (1981) report on in their review analyse at least a part of the total model.

The two examples show that the path diagram is helpful for the clarification of complex theories. Eventually the model may be be confronted with and tested against appropriate data. As long as the path diagram is used only for theoretical consideratons, and as long as no estimation is intended, the difference between the ML and the PLS method is unimportant. But when it comes to the estimation, differences between ML and PLS come into the picture, some of them related to the statistical foundations of the methods, some of them related to the existing programs.

The systematic differences will be discussed in the next three sections. The fifth section shows several applications and demonstrates the range of unique applications of PLS.

2. A two-block model, estimated by PLS and ML.

One essential difference between the Maximum Likelihood (ML) method and the Partial Least Squares (PLS) method of parameter estimation for LVP models can be demonstrated by means of a very small model which contains only two latent variables, each one indicated by a set of three manifest variables. PLS versions of this model are the well-known Canonical Correlation (CC) model and Tucker's (1959) interbattery factor (IBF) model. Everybody who has toyed around with LISREL and his text-book knowledge on multivariate statistics has found that the LISREL, CC and IBF results are different. A small artifical correlation matrix will suffice to demonstrate the main aspects of the difference. A small real world data example will be given later.

Suppose a sample correlation matrix S=cor(x,y) is given for two blocks of standardized variables x_j j=1..J and y_k k=1..K with J=K=3,

Eq.01 $$S = 1/100 * \begin{pmatrix} 100 & & & & & \\ 10 & 100 & & & & \\ 10 & 10 & 100 & & & \\ 09 & 09 & 09 & 100 & & \\ 09 & 09 & 09 & 10 & 100 & \\ 09 & 09 & 09 & 10 & 10 & 100 \end{pmatrix}.$$

All correlation coefficients are very small, the correlations between blocks (s_{jk}=0.09) being a little bit smaller than the correlations within blocks ($s_{jj'}=s_{kk'}$=0.10). Depending on sample size, in most cases it would make no sense at all to analyse the association between the two blocks, because the correlation cannot securely be distinguished from a zero correlation.

However, let us assume a model with a latent variable ξ underlying the manifest variables x_j, and a latent variables η for the MVs y_k. The parameter of interest is the correlations coefficient $\rho=cor(\xi;\eta)$. The first estimation method to be studied is the canonical correlation analysis which is a least squares (LS) method and gives the estimate

Eq.02 $r_{CC} = 0.225.$

The second method uses LISREL which is a maximum likelihood (ML) method and gives the estimate

Eq.03 $r_{ML} = 0.900.$

The LISREL result seems to be unbelievable and extremely exaggerated, because the estimated association $r_{\xi\eta}$=0.900 is ten times as high as the correlation s_{xy}=0.090 between the observed variables. This difference is to be explained.

The LISREL model: We assume the simple model as shown in Table 3, where all variables are standardized, the loading coefficient (pattern coefficient) is denoted by π, the path coefficient between the two LVs ξ and η by β, and the variance of the outer residual variables by θ^2. From the simple rules of path analysis (cf. Heise 1975) it follows that the covariance/correlation matrix of the manifest variables has the theoretical structure

Eq.04
$$\Sigma = \begin{pmatrix} \pi\pi+\theta^2 & & & & & \\ \pi\pi & \pi\pi+\theta^2 & & & & \\ \pi\pi & \pi\pi & \pi\pi+\theta^2 & & & \\ \pi\beta\pi & \pi\beta\pi & \pi\beta\pi & \pi\pi+\theta^2 & & \\ \pi\beta\pi & \pi\beta\pi & \pi\beta\pi & \pi\pi & \pi\pi+\theta^2 & \\ \pi\beta\pi & \pi\beta\pi & \pi\beta\pi & \pi\pi & \pi\pi & \pi\pi+\theta^2 \end{pmatrix}$$

Given the sample correlation matrix S in Eq.01, the only solution which gives a perfect fit to Σ as defined in Eq.04 has the parameters

Eq.05 $\pi = \sqrt{0.10}, \quad \beta=0.90, \quad \theta^2=0.90$

The path coefficient β can be found in this simple case by

Eq.06 $\beta = \pi\beta\pi/\pi\pi = 0.09/0.10 = 0.90$

the ratio of between-block correlation over within-block correlation. (Question: What happens when the between-block correlation is 0.11 and the within-block correlation is 0.10?)

The Canonical Correlation model: The CC model prescribes to form estimates for the latent variables as standardized weighted aggregates of the

manifest variables

Eq.07 $\hat{\xi} = \sum_j v_j x_j$

Eq.08 $\hat{\eta} = \sum_k w_k y_k$

such that the correlation between these two estimates is maximal ·

Eq.09 $\rho = \text{cor}(\hat{\xi};\hat{\eta}) = \text{max.}$

Ignoring the standardization for a moment, we chose the weight for the sample correlation matrix S in Eq.01 to be all equal, $v_j = w_k = 1$. Then $\text{var}(\hat{\xi}) = \sum_j \sum_{j'} 1*1*s_{jj'} = 3.6$, $\text{var}(\hat{\eta}) = \sum_k \sum_{k'} 1*1*s_{kk'} = 3.6$, and $\text{cov}(\hat{\xi};\hat{\eta}) = \sum_j \sum_k 1*1*s_{jk} = 9*0.09 = 0.81$.

Eq.10 $$\text{cov}(\hat{\xi}, \hat{\eta}) = \begin{pmatrix} 3.60 & 0.81 \\ 0.81 & 3.60 \end{pmatrix},$$

and the regression coefficient (for standardized variables identical to the correlation coefficient) is

Eq.11 $b(CC) = \sum_j \sum_k s_{jk} / \sum_j \sum_{j'} s_{jj'} = 0.81/3.60 = 0.255.$

The first essential difference between the ML and LS method lies in the prescription, that in the LS method the LV is to be estimated as a linear combination of its indicators. By this requirement, the LV is closely related to the observed and measured data. The multiple correlation of the LV on its MVs is perfect in the LS method, but not so in the ML method. We re-write the correlation matrix for the x-variables, together with their correlations with the estimated LV $\hat{\xi}$ (the so-called LV-structure coefficients).

Eq.12 $$\text{cov}(x, \hat{\xi}) = 1/100 * \begin{pmatrix} 100 & & & \text{PLS} \\ 10 & 100 & & \\ 10 & 10 & 100 & \\ 63 & 63 & 63 & 100 \end{pmatrix}, \quad \text{or} \quad \begin{pmatrix} 100 & & & \text{ML} \\ 10 & 100 & & \\ 10 & 10 & 100 & \\ 32 & 32 & 32 & 100 \end{pmatrix}$$

The LV-structure coefficient for the ML estimate is $\sqrt{0.10}=0.32$, for the LS estimate = $\sum s_{jj'} / \hat{\sigma}_\xi = 1.20/ 3.60 = 0.63$. The squared multiple correlation $R^2(\hat{\xi},x)$ is for the LS case $R^2=1$ by prescription, for the ML case $R^2=0.25$. Thus the LV in the ML solution is far from being palpable, more the less the

correlation r=0.90 between these untouchable variables. The LS estimates of the LVs are tied completely to the observations, and the estimate of the correlation between the LVs is very conservative, on the secure side.

Conclusions: The mathematical considerations on the model show very clearly that the ML estimate is correct. Thus it depends on the researcher whether he "believes" more in his data or in his model. If he is shure that the model is correct, and that his MVs x_j and y_k are valid indicators of the MVs ξ and η, and that is only due to a practical reason (not a theoretical reason) that no more MVs are available, then he may accept the ML estimate as the figure he is interested in. However, he should be aware that his measure on the association between the two unobserved constructs ξ and η are very far from being observable. And if the sample correlation matrix in Eq.01 is not gathered with sufficient precision, and if he gets e.g. $s_{jk}=0.11$ instead of $s_{jk}=0.09$, than the correct LISREL estimate will turn out to be r=1.10.

The LS estimate is on the conservative side, because it overestimates the loadings, and thus links the solution more closely to the observations, and it underestimates the relationshiops between the LVs. It can never happen that the correlation $r(\hat{\xi}; \hat{\eta})$ exceeds unity, or that the residual variance is negative (or other Heywood cases), because the LV correlation is the correlation between linear combinations.

Table 4 shows the real world data example that led to the construction of the more extreme artificial data in Eq.01. This example is taken from Weber (1982) who did not stress so much on the very high correlation cofficient he found by LISREL. The comparison to the CC and IBF estimates indicates that the estimate can depend highly on the statistical assumption introduced only to facilitate the estimation.

LS methods are more close to the data, more explorative, more data analytic. ML methods are more close to the model, more confirmative, more model analytic.

3. The statistical properties: consistency.

In the previous section, it was referred to the limited number of manifest variables available as indicators for a construct variable. Is it true that the number is limited only due to practical reasons? This is questionable, especially in the human sciences. For example, is it possible to find arbitrary many indicators for the socio-economic status of the family of a student, after education and occupation of the parents are taken as indicators? If we include the income of the father into the set of indicators, the meaning of the LV is likely to change a little bit towards a stress on the socio-economic status of the father. If we include the number of books in the household as indicator, the LV will tend to become a socio-cultural indicator. In general, it is for theoretical reasons that the number of indicators is limited.

Experience tells as that no empirical correlation matrix looks like Eq.01. Each indicator behaves a little bit different; the first one is correlated to the second one a little bit stronger than to the third one, and this shows that the first and second indicator are influenced by a doublet factor. And if the first three indicators are related to each other a little bit stronger than to a fourth indicator, these three can be understood as to be influenced by a triplet facter. With each new indicator introduced in the set of indicators, possibly a whole bunch of factors is introduced. A careful selection of indicators may help to reduce the influence of all factors others than the unique factors and the common factor, but cannot eliminate them totally.

To put it otherwise: if only very few indicators are available, it is nearly impossible to assert the assumption that the indicators measure nothing else than the indended construct, and that no doublet, triplet, etc. factors are included. If there are many indicators, the unidimensionality of the MVs can be proved by the data. The ML method and the PLS method take different

conclusions from this fact.

PLS: If there is not more than one indicator, the PLS estimate for the LV is identical to the MV, which is so-to-say a non-estimate. If there are three indicators (cf. example in Sc.2), the estimate of the LV is too close to the data, and the relations between the LVs are underestimated. The more indicators are used, the closer come the estimates of the inner relationships to the "true" parameters.

This can be demonstrated by Eq.10. Suppose the number of MVs in Eq.01 is increased from J=K=3 to J=K=6, and the correlation within the blocks are all $s_{jj'} = s_{kk'} = 0.10$ and between blocks $s_{jk} = 0.09$, as unrealistic as it may be. Then, with $v_j=w_j=1$, we get as covariance of the LVs

Eq.13 $$\mathrm{cov}(\hat{\xi}; \hat{\eta}) = \begin{pmatrix} 12.00 & 3.24 \\ 3.24 & 12.00 \end{pmatrix}$$

whence it follows that $r=r(\hat{\xi};\hat{\eta}) = 3.24/12.00 = 0.27$. With J=K=12 indicators, we get r = 12.96/25.20 = 0.51, and with J=K=24 we get r=0.68.

With increasing number of indicators, the estimate of the correlation between the LVs comes closer to the true value. This is what is said in the theorem on consistency at large (Wold 1982): The PLS estimates approach the true values as the number of cases in the sample increases (i.e. the sample correlation matrix becomes a better estimate of the population correlation matrix), and as the number of indicators increases. The substantive interpretation and justification of this theorem is obvious: The more the unidimensionality of the a block of indicators can be proved by the data, i.e. the more the model assumptions are justified by the data, the more it makes sense at all to estimate the parameters of the model, and the more the parameter estimates become consistent. Conversely: if the model assumptions cannot be asserted by the data, it is better to remain closely to the data than to estimate the parameters of the model.

ML: The ML estimates are always consistent, no matter how poor the

indicators and the data are. If the model is totally wrong, the LISREL iterative estimation is not likely to converge.

4. Covariance structure, data structure.

The differences of the PLS and ML estimation technique for LVP models are discussed in Sc.2 with respect to the results, in Sc.3 with respect to the statistical properties of the results. In this section we will look into the estimation criterion itselves. The difference between PLS and ML will be traced back to a difference between covariance structure analysis and data structure analysis. The difference will be explained using a special case, the oblique factor model.

Data structure: Let y denote the (KxN) data matrix, ε the (KxN) matrix of outer residual scores, η the (JxN) matrix of factor scores, υ the (KxN) matrix of inner residual scores, Π the (JxK) pattern matrix, Ω the (JxK) matrix of factor weights. Then the model for the data is written as

Eq.14 $\eta_{jn} = \sum_k \omega_{kj}\, y_{kn} + \upsilon_{jn}$

Eq.15 $y_{kn} = \sum_j \pi_{kj}\, \eta_{jn} + \varepsilon_{kn}$

Notice that the scores on the LV are written as a linear functions of the MVs and a residual variable which will be assumed to be zero when it comes to the estimation. Eq.15 is the conventional factor model. The Ordinary Least Squares (OLS) estimation criterion is data oriented and minimizes the discrepancies between the measured values y_{kn} and the estimated values $\hat{y}_{kn}$

Eq.16 $OLS = 1/N \sum_n \sum_k (y_{kn} - \hat{y}_{kn})^2$

Covariance structure: Let Σ denote the (JxJ) covariance matrix of y, Φ the (KxK) correlation matrix of the LVs, and Θ the (KxK) residual covariance matrix. Then the model equation for the covariance matrix can be written as

Eq.17 $\Sigma = \Sigma(\Pi, \Phi, \Theta, cov(\eta;\varepsilon)) = \Pi.\Phi.\Pi' + \Theta + cov(\eta;\varepsilon) + cov(\varepsilon;\eta)$

The Unweighted Least-Squares (ULS) criterion (as well as the Generalized Least-Squares (GLS) and the Maximum Likelihood (ML) criterion) is covariance

oriented and minimizes the discrepancies between the empirical covariance $s_{kk'}$ and the expected (deduced, reproduced) covariance $\hat{\sigma}_{kk'}$ (cf. Joereskog 1978)

Eq.18 $\quad ULS = \sum_k \sum_{k'} (s_{kk'} - \hat{\sigma}_{kk'})^2 = \text{trace}(S - \hat{\Sigma})^2 = \min$

Eq.19 $\quad GLS = \text{trace}(I - S^{-1}\hat{\Sigma}) = \min.$

Eq.20 $\quad ML = \text{trace}(\hat{\Sigma}^{-1} S) - \log|\hat{\Sigma}^{-1} S| = \min.$

Comparison of Eq.16 and Eq.18 shows that the OLS criterion is applied to the data matrix, and that the ULS,GLS,ML criteria are applied to the covariance matrux. The OLS critions can be re-formulated as a criterion

Eq.21 $\quad OLS = \text{trace}(S-\hat{\Sigma}) = \min$

applied to the diagonal of the covariance matrix ("Maximize the explained variance"). In most applications of covariance structure model Eq.17 it is assumed that $\text{cov}(\eta;\varepsilon)=0$ and Θ=diagonal, whence it follows that a perfect fit to the diagonal of S is allways possible, and that the remaining parameters are to be estimated such that the off-diagonal elements of S are optimally reproduced ("Maximize the explained covariance").

Covariance structure analysis is based upon the assumption that all relevant information is contained in the covariance matrix, and that the data have no excess information. This assumption is tenable when the data is considered to be only a sample taken from the population of interest. Data structure analysis, as opposed, does not go beyond the data at hand, and the estimation is done without reference to a population.

That the covariance matrix contains all the information about the joint distribution of the variables, is true once it is ashured that the relationships between the variables are linear as stated in Eq.15. The linearity assumption is crucial for both estimation methods, the LS and the ML method. If there is theoretical or empirical evidence that the relationships are nonlinear, the model builder has to re-specify the model and to include squared, logarithmic, or product variables in order to account for nonlineari-

ties. It is even possible to specify the model in a way that for different subsamples different parameters are admitted (Lohmoeller & Wold 1982).

Is it a big step from the linearity assumption to the multinormality assumption? Once it is assumed that all variables are jointly multivariate normal distributed, it follows immediately that all relationships between all variables involved are linear. Dijkstra (1981) argues that only very mild additional assumptions are required, and the linearity assumption is asymptotically equivalent to the multinormality assumption. However, Dijkstra considers only the very special case where the outer residual covariance matrix is assumed as diagonal.

Summing up the results about data structure analysis by least squares methods:

(i) The data structure analysis includes the LV scores (or the LV weights) in to the basic set of unknowns, but not the residual covariances. The IVs are deliberately estimated as weighted aggregates of the MVs. No assumptions are made about the residual covariances.

(ii) Data structure analysis makes no explicite reference to a population, and the distributional assumption is relatively soft: Only linearity, no multinormality.

(iii) The estimation criterion request a minimization of the residual variances, not of the residual covariances.

(iv) As a result, the measured data points are reconstructed (predicted) as well as possible by the estimated LV scores and other parameters, in a least-squares sense. The ML method aims at optimal parameters. You can not have both, optimal prediction and optimal parameters, as a theorem by H.Wold (1981) states.

5. Some selected applications of PLS.

In this section, several extreme applications of the PLS estimation to LVP models will be reported on.

Very much variables: Bartl, Unverdorben and Lohmoeller (1981) describe a model on alcohol and suicide problems of German army draftees. The model includes 102 MVs and 23 LVs. The estimation by LVPLS took 200 seconds. An attempt to estimate the model by LISREL was given up after LISREL had done four iteration in 30 minutes. Models of similar size are published by Gattringer (1980).

Very few cases: Wold (1980) analysed data on economic sanctions against bad behaving countries. Between 1933 and 1965 only 10 countries were punished by economic boycotts, but 27 variables describing the relationships of the sanctioner and the sanctioned country are available. Wold demonstrates by means of the blindfolding technique (similar to jack-knife technique) that PLS collects relevant information into LVs, even when there are three times more variables than cases.

Few cases, much variables: Schulte (1981) anlyses data on a psychomedical group training during a cure. He has 100 cases and 100 variables in his longitudinal design. More exactly, there are 100 cases in the beginning, filling out the first questionaire. But the last questionaire, 12 weeks after the end of the cure, was returned only by 30 patients. Schulte's model includes 96 MVs and 26 LVs. The estimation procedure used the missing-data correlation matrix, which were singular. Even the LV-correlation matrix were singular, and strong restrictions on the path model were to be introduced in order to avoid a selection of a singular set of predictors.

Categorical variables: Wold and Bertholet (1981, 1982) show that biviarate and multivariate contingency tables can be analysed by LVP models and PLS estimation. Interactions of any order can be introduced.

Mixed-scale variables: Lohmoeller and Wold (1982) show how categorical and interval-scaled variables can be included into one model. The estimation procedure uses a product moment matrix which can include correlations, covariances, contingency tables, and group-specific averages. It is possible to provide different loading matrices for different groups.

Multiple level LVs: Hierarchical factor and path models with latent variables are treated by Bentler (1982) with ML methods, and by Noonan and Wold (1982) with PLS methods.

LVPLS and LISREL: The program LVPLS can be used as a forerunner for LISREL. LVPLS generates a complete input deck for LISREL, the PLS estimates included as starting values. Typically, LISREL uses 5 times as much computer time as LVPLS, even with the reasonable LVPLS starting values (Lohmoeller 1981a).

LVPLS and RAM: McArdle (1982) used LVPLS to get starting values for his version of the LVP analysis he calls RAM analysis. The LVPLS estimation for a model with 28 MVs took about 100 cpu seconds. The ML estimates were obtained from his MULTITAB program in approximately 7500 cpu seconds.

6. Conclusions.

LVP modeling, even if it is not intended to end in a statistical analysis, has found broad application in social sciences. It is clarifying toughts, ideas and hypothesis, and brings them together into a closed picture of a theory.

If the LVP model is to be confronted with data, it depends sometimes on the sheer number of variables and cases, wether a ML estimation is possible or not. If the conditions for the application of ML methods are violated (missing multinormality, singular covariance matrix), the PLS method may be the only possible one.

Theoretical knowledge is sometimes scarce, exspecially in interdisciplinary studies. Here it would make no sense to aim at consistent estimator of parameters of a model, because theoretical knowledge is to small to construct a reasonable model. The PLS methods provides a continuum from less to more restrictive formulations of the model, from data analysis to model analysis, from data exploration to hypothesis and model testing.

References.

Bartl,P., K.Unverdorben, J.B.Lohmoeller (1981) Soziale Probleme im Grundwehrdienst: Eine Pfadanalyse zu Alkohol- und Suicidproblemen. Forschungsbericht 81.01 Fachbereich Paedagogik, Hochschule der Bundeswehr Muenchen.

Bentler,P.M. (1980a) The interdependence of theory, methodology, and empirical data: Causal modeling as an approach to construct validation. In D.B. Kandel (Ed.), Longitudinal research on drug use: Empirical findings and methodological issues. New York: Wiley.

Bentler,P.M. (1980b) Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology, 31:419-456.

Bentler,P.M. (1982) Linear systems with multiple levels and types of latent variables. In K.G.Joereskog and H.Wold (Eds.), Systems under indirect observation: Causality, structure, prediction. Amsterdam: North Holland (in press).

Bentler,P.M & D.G.Weeks (1980) Linear structural equations with latent variables. Psychometrika, 45:289-308.

Dijkstra,T.K. (1981) Latent variables in linear stochastic models: Reflections on Maximum Likelihood and Partial Least Squares Methods. Unpublished Diss Uni Groningen, The Netherlands.

Gattringer,H. (1980) Die Entwicklung des Leistungsmotivs und dessen Bedeutung fuer die Schulleistungsprognose. Diss. University Salzburg, Austria.

Glasman,N.S. & I.Biniaminov (1981) Input-output analyses of schools. Review of Educational Research, 51:509-539.

Heise,D. (1975) Causal analysis. New York: Wiley Interscience.

Joereskog,K.G. (1973) A general method for estimating a linear structural equation system. Pp 85-112 in A.S.Goldberger & O.D.Duncan (Eds.), Struc-

tural equation models in the social sciences. New York: Seminar Press.

Joereskog,K.G. (1978) Structural analysis of covariance and correlation matrices. Psychometrika, 43:443-447.

Joereskog,K.G. & D.Soerbom (1981) LISREL V: Analysis of linear structural relationships by maximum likelihood and least squares methods. Research report 81-8, Department of Statistics, University of Uppsala, Sweden.

Joereskog,K.G. & H.Wold (1982 Eds.) Systems under indirect observation: Causality, structure, prediction. Amsterdam: North Holland (in press).

Lohmoeller,J.B. (1979) Estimating parameters of linear structural relation models under partial least-squares criteria. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April 1979 (ERIC no. ED 178 600).

Lohmoeller,J.B. (1981a) Pfadmodelle mit latenten Variablem: LVPLS ist eine leistungsfaehige Alternative zu LISREL. Forschungsbericht 81.02 Fachbereich Paedagogik, Hochschule der Bundeswehr Muenchen.

Lohmoeller,J.B. (1981b) LVPLS 1.6 program manual: Latent variables path analysis with partial least squares estimation. Forschungsbericht 81.04 Fachbereich Paedagogik, Hochschule der Bundeswehr Muenchen.

Lohmoeller,J.B. & H.Wold (1982) Path models for latent variables indirectly observed by categorical and interval-scaled variables: An extension of Partial Least Squares (PLS) Soft Modeling. Paper presented at the Symposium on Cultural Indicators for the Comparative Study of Culture, Oesterreichische Akademie der Wissenschaften, Vienna, Feb.16-19, 1982.

McArdle,J.J. (1982) Structural equation modeling of an individual system: Preliminary results from "A case of episodic alcoholism." Unpublished paper, draft Jan.1982, Department of Psychology, University of Denver.

McArdle,J.J. & R.P.McDonald (1981) A simple algebraic representation for structural equation models. Unpublished paper, draft Oct.1981, Department of Psychology, University of Denver.

Noonan,R. & H.Wold (1982) Evaluation of educational systems using Partial Least Squares. In B.H.Choppin & T.N.Postlethwaite (Eds.), Journal of Educational Systems Evaluation. London: Pergamon (forthcoming).

Schulte,Ch. (1981) Ueberlegungen zur Genese und Aufrechterhaltung psychosomatischer Stoerungen: Das verhaltensmedizinische Gruppentraining, seine Entwicklung und Evaluation in der klinischen Rehabilitation (Diss Uni Tuebingen). Garmisch-Partenkirchen: Kroll-Verlag.

Shavelson,R.J. & P.Stern (1981) Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. Review of Educational Research, 51:445-498.

Weber,R.P. (1982) Content-analytic cultural indicators. Paper presented at the Symposium on Cultural Indicators for the Comparative Study of Culture, Oesterreichische Akademie der Wissenschaften, Vienna, Austria, Feb.16-19, 1982.

Wold,H. (1966a) Estimation of principal components and related models by iteratative least squares. In P.R.Krishnaiah (Ed.), Multivariate analysis. New York: Academic.

Wold,H. (1966b) Nonlinear estimation by iterative least squares procedures. In F.N.David (Ed.), Festschrift for J.Neyman: Research papers in statistics. London: Wiley.

Wold,H. (1975 Ed.) Modelling in complex situations with soft information. Third World Congress of Econometrics, Toronto, Canada, Aug.21-26, 1975.

Wold,H. (1975b) From hard to soft modelling. In H. Wold (Ed.), Modelling in complex situations with soft information. Third World Congress of Econometrics, Toronto, Canada, Aug.21-26, 1975.

Wold,H. (1979) Model construction and evaluation when theoretical knowledge is scarce: An example of the use of partial least squares. Cahier 79.06 Département d'économétrie, université de Geneve, Switzerland.

Wold,H. (1980) Factors influencing the outcome of economic sanctions: An

application of soft modelling. Paper presented at Fourth World Congress of Economic Society, Aix-en-Provence, France, Aug.28 - Sep.02, 1980.

Wold,H. (1981a) Model construction and evaluation when theoretical knowledge is scarce: On the theory and application of Partial Least Squares. In J.Kmenta & J.Ramsey (Eds.), Model evaluation in econometrics. New York: Academic Press.

Wold,H. (1981b) Systems under indirect observations using Soft Modelling. Working Paper No. 48, Economics Department, Case Western Reserve University, Cleveland, Ohio.

Wold,H. (1981c) Comments on the papers by J.B.Grubman & E.T.H.Shih, J.W.Downs & D.M.Rocke, R.Litterman, and C.W.J.Granger, comments integrated with a briefing of PLS (Partial Least Squares) soft modeling. Conference on Applied Time Series Analysis of Economic Data, co-sponsored by the American Statistical Association, U.S. Department of Commerce, Bureau of the Census, and the National Bureau of Economic Research, Pentagon City, Arlington, Virginia, Oct.13-15, 1981.

Wold,H. (1982) Soft modeling: The basic design, and some extensions. In K.G. Joereskog & H.Wold (Eds), Systems under indirect observation: Causality, structure, prediction. Amsterdam: North Holland (in press).

Wold,H. & J.L.Bertholet (1981) The PLS (Partial Least Squares) approach to multidimensional contingency tables. In A. Rizzi (Ed.) Transaction of the International meeting on Multidimensional Contingency Tables, University of Rome, June 25-26, 1981 (Special issue of Metron, to appear).

Wold,H. & J.L.Bertholet (1982) The PLS (Partial Least Squares) approach to multidimensional contingency tables, II. Paper presented at the Annual Meeting of the American Educational Research Association, March 19 - 23, 1982, New York.

Figure 1: Overview of the domain of research on teachers' judgments, decisions and behavior (from Shavelson and Stern 1981)

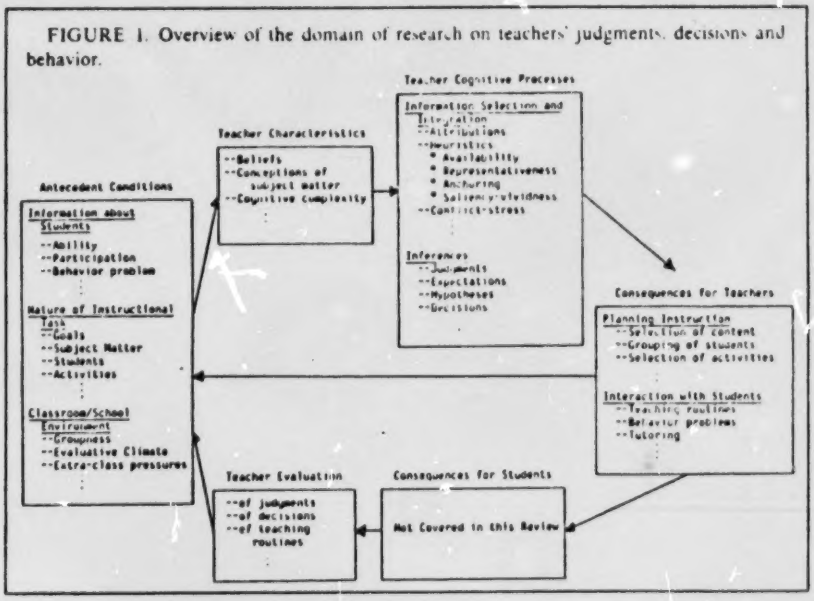

Figure 2: A suggested structural model of school input and output variables (from Glasman and Biniaminov 1981).

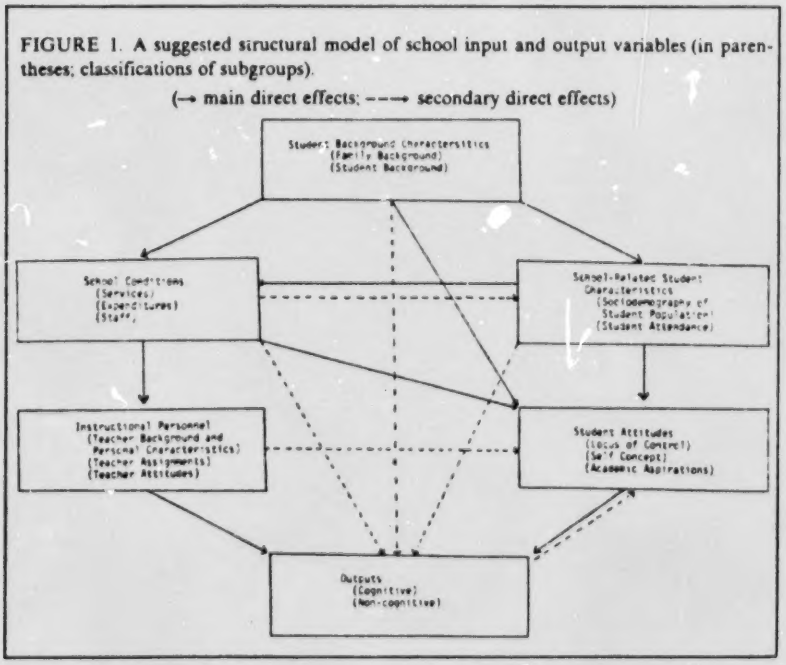

Figure 3: A two-block model.

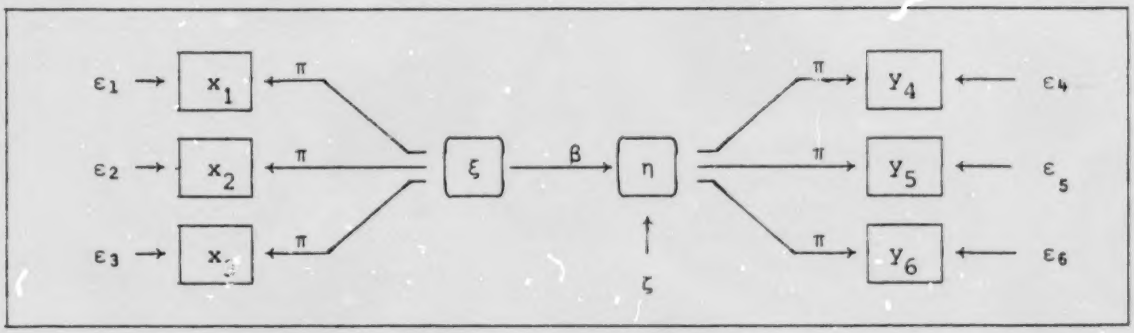

Table 4: Association between German economic development (as measured by Berlin grain prices) and concern of Wealth in the throne speeches of the Kaiser: LVP analyses with LISREL and LVPLS (Data from Weber 1982).

Variable		PLS mode A IBF model		PLS mode B CC model		LISREL	
Rye price (2nd diff.),	p=	23.92	0	24.88	0	21.90	0
Wheat price (2nd diff.)		20.65	0	18.70	0	16.20	0
Wealth - others		0	1.89	0	1.23	0	1.00
Wealth - transactions		0	.53	0	.84	0	.64
Wealth - participants		0	.31	0	.37	0	.30
Economic performance	r =	1.00		1.00		1.00	
Wealth concerns		-.47	1.00	-.59	1.00	-.85	1.00

Economic performance: Indicators are the prices at the Berlin wholsale grain market form 1871 to 1912, more exactly the second differences, i.e. the speed of price changes. Scale unit: Mark.
Wealth concerns: Content analytic indicators, extracted from the speeches of the German Kaiser, 1871 - 1912. Scale unit: number of words per 1000.
Correlation: "As the economy improves there is less concern with wealth, while a declining economy is associated with increasing wealth concerns".